

Conference Abstract

How Agricultural Researchers Share their Data: a Landscape Inventory

Cynthia Parr[‡], Erin Antognoli[§], Jonathan Sears[§]

[‡] National Agricultural Library, USDA, Beltsville, MD, United States of America

[§] LAC Federal at National Agricultural Library, Beltsville, United States of America

Corresponding author: Cynthia Parr (cynthia.parr@ars.usda.gov)

Received: 18 Aug 2017 | Published: 20 Aug 2017

Citation: Parr C, Antognoli E, Sears J (2017) How Agricultural Researchers Share their Data: a Landscape Inventory. Proceedings of TDWG 1: e20434. <https://doi.org/10.3897/tdwgproceedings.1.20434>

Abstract

The United States Agricultural Research Service (ARS) recently declared a grand challenge: Transform agriculture to deliver a 20% increase in quality*1 food availability with 20% lower environmental impact by 2025. Addressing this challenge requires a sea change in how it conducts agricultural research. Not only will teams need to be multi-disciplinary, as they begin to pursue big data and data-intensive approaches, they will need to find effective ways to share their diverse kinds of data with each other, with other research teams, with members of farming and business communities, and with policy-makers. Biodiversity is a key component of food production (crop and livestock species, for example, and the pollinators and microbes they depend on) and the impact that food production (including reduction of pest and pathogen species) has on the environment (species richness, invasive species, and ecosystem services, for example). It is currently unclear how much biodiversity data relevant to agriculture is being made available, and if so where it is. These questions are part of a general need to understand how our pilot platform for USDA-funded data cataloging and publication, the Ag Data Commons <https://data.nal.usda.gov>, can best support grand challenge research. It will also help agricultural librarians assist their researchers in data management and publication. Therefore we conducted an extensive inventory of the options available to researchers both for finding data and sharing data related to the broader areas of agricultural research. We present the general results for agriculture overall, then explore the agrobiodiversity sector specifically. We found 230 active and publicly available agriculture-specific databases and repositories,

only 16.6% of which accept submissions outside their institution, consortium or projects, and most of which are not using or not relevant to TDWG standards such as Darwin Core. The use of taxonomic identifiers is also not standardized. While 73 more general repositories (including the Global Biodiversity Information Facility, GBIF) have easily discoverable agricultural data, in many cases the amounts are currently much smaller than one might expect given vast investments in agricultural research. We reviewed the total number of datasets returned by seven agriculture-related search terms, as well as the percent of the total repository each term represented. Only twenty-five (34.2%) of the general repositories returned over 500 results from at least one agricultural search term. Only ten repositories (13.7%) returned 5% or more of their collection with any of these agricultural search terms. Of the top 50 journals where USDA researchers published in 2016, 40 (80%) host supplemental datasets and most state that supplemental material is published as submitted and will not be edited. Thirty (60%) either require or strongly encourage authors to deposit supporting data in public repositories, with 21 (42%) recommending discipline-specific repositories (four journals name GBIF, for example). Only one journal recommended metadata standards according to type of data. Future work should include an assessment of how many of these databases and repositories have machine-readable data dictionaries, which could be used to more effectively discover agriculturally-relevant data and to foster meaningful data integration. Future work should also explore how mining the Biodiversity Heritage Library and other sources can increase the availability of machine-readable legacy agrobiodiversity data.

Keywords

agrobiodiversity, data repositories, public access, open data, data publication

Presenting author

Erin Antognoli

Endnotes

*1 nutritious and safe